



Blatt 2

Aufgabe 2.1 Berechnen Sie den Erwartungswert und die Varianz der geometrischen Verteilung.

Aufgabe 2.2 Gegeben seien 4 Münzen M_1, M_2, M_3, M_4 , die nacheinander jeweils mehrfach geworfen werden. Die Münzen sind zum Teil manipuliert und haben daher verschiedene Wahrscheinlichkeiten p_{M_i} dafür, dass bei einem Wurf Kopf fällt:

$$p_{M_1} = 0.4 \quad , \quad p_{M_2} = 0.5 \quad , \quad p_{M_3} = 0.3 \quad , \quad p_{M_4} = 0.5$$

- Wie gross ist die Wahrscheinlichkeit, dass eine einzelne Wurfsequenz S_N geworfen wird? Eine Sequenz S_N sei gegeben durch 4 Werte, die jeweils die Anzahl von Kopfeignissen für die einzelnen Münzen bei N Würfungen angeben. $S_{10} = [3, 5, 6, 8]$ bedeutet also, dass bei 10 Würfungen mit jeder Münze bei M_1 3 mal Kopf fiel, usw.
- Schreiben Sie eine Funktion, die Ihnen für eine gegebene Wurfsequenz S_N deren Wahrscheinlichkeit errechnet. Rückgabewert soll dabei die logarithmierte Wahrscheinlichkeit sein, die oftmals in der Praxis Verwendung findet. Was ist bei der numerischen Berechnung der Wahrscheinlichkeiten zu berücksichtigen? Testen Sie Ihre Funktion für verschiedene Sequenzen mit verschieden großen Werten für N .

Aufgabe 2.3 Wir betrachten die idealisierte shot-gun Fragmentierung einer Sequenz der Länge G mit insgesamt N Fragmenten jeweils der Länge L . Das linke Ende jedes Fragments liege (wiederum vergrößert) gleichverteilt auf einer der Positionen im Intervall $[1, G]$.

- Für welche Anzahl N der Fragmente wird die erwartete Anzahl $C = e^{-a}N$ maximal? Wir halten dabei G und L fest, $a = \frac{LN}{G}$ sei die coverage. Vergleichen Sie dieses Ergebnis mit den Beispielzahlen der Vorlesung.
- Für $|x|$ klein gilt folgende Approximation: $e^x \approx 1 + x$. Nutzen Sie diese Approximation für die coverage a in der Abschätzung über die erwartete contig-Länge $L \frac{e^a - 1}{a}$. Interpretieren Sie das Ergebnis.
- Simulieren Sie diese shotgun-Fragmentierung für verschiedene Werte von G, L und N und vergleichen Sie die Resultate mit den Vorhersagen. Ändern die nun die Simulation, indem die Länge L der Fragmente gleichverteilt aus dem Intervall $[400, 600]$ gezogen wird und vergleichen Sie.